A Brief Introduction to Bayesian Networks

AIMA 14.1-14.3

CIS 391—Intro to Artificial Intelligence

(LDA slides from Lyle Ungar from slides by Jonathan Huang

(jch1@cs.cmu.edu))
Bayesian networks

- A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions

- Syntax:
  - A set of nodes, one per random variable
  - A set of directed edges (link ≈ "directly influences"), yielding a directed acyclic graph
  - A conditional distribution for each node given its parents: \( P(X_i \mid \text{Parents}(X_i)) \)

- In the simplest case, conditional distribution represented as a conditional probability table (CPT) giving the distribution over \( X_i \) for each combination of parent values
The Required-By-Law Burglar Example

I'm at work, and my neighbor John calls, but my neighbor Mary doesn't call. They usually call if my burglar alarm goes off, and occasionally otherwise. Sometimes the alarm is set off by minor earthquakes. Is there a burglar?

Variables: Burglary, Earthquake, Alarm, JohnCalls, MaryCalls

Network topology reflects causal knowledge:
1. A burglar can cause the alarm to go off
2. An earthquake can cause the alarm to go off
3. The alarm can cause Mary to call
4. The alarm can cause John to call
Belief network for example

Network topology reflects "causal" knowledge:
1. A burglar can cause the alarm to go off
2. An earthquake can cause the alarm to go off
3. The alarm can cause Mary to call (with prob \( \approx .70 \))
4. The alarm can cause John to call (with prob \( \approx .90 \))
Network topology reflects "causal" knowledge:

1. A burglar can cause the alarm to go off (with probability \( \approx .94 \))
2. An earthquake can cause the alarm to go off (with prob \( \approx .29 \))
3. The alarm can cause Mary to call
4. The alarm can cause John to call
Belief network for example

Unconditional probabilities reflect priors of those events:

1. The probability of a burglary (per day?) is .001
2. The probability of an earthquake is .002
Belief network with all (conditional) probability tables

Makes it easy to compute answers to questions like:

- How likely is John to call if there’s an earthquake?
- How likely is Mary to call on a random day?
Semantics

- *Local* semantics give rise to *global* semantics
- Local semantics: *given its parents, each node is conditionally independent of its other ancestors*

**REVIEW** (clarity is *crucial*):
- A and B are *conditionally independent given C* iff
  - \( P(A \mid B, C) = P(A \mid C) \)
  - Alternatively, \( P(B \mid A, C) = P(B \mid C) \)
  - Alternatively, \( P(A \land B \mid C) = P(A \mid C) \ast P(B \mid C) \)
- Review: Toothache (T), Spot in Xray (X), Cavity (C)
  - None of these propositions are independent of one other
  - But *T & X are conditionally independent given C*
Semantics

- Local semantics give rise to global semantics
- Local semantics: given its parents, each node is conditionally independent of its other ancestors
- Global semantics (why?):

\[ P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \text{Parents}(X_i)) \]

\[ P(J, M, A, \neg B, \neg E) = P(\neg B)P(\neg E)P(A | \neg B, \neg E)P(J | A)P(M | A) \]

Generalizes Naïve Bayes
Naïve Bayes as a Bayesian Network

\[ P(Flu, X_1, \ldots, X_5) = P(Flu) \cdot P(X_1 \mid Flu) \cdot \cdots \cdot P(X_5 \mid Flu) \]
Belief Network Construction Algorithm

1. Choose some ordering of variables $X_1, \ldots, X_n$

2. For $i = 1$ to $n$
   a. add $X_i$ to the network
   b. select parents from $X_1, \ldots, X_{i-1}$ such that

   $$P(X_i|\text{Parents}(X_i)) = P(X_i|X_1, \ldots, X_{i-1})$$

3. Yields the smallest networks when \textit{causers} (\textit{diseases}) precede consequences (\textit{symptoms})
Suppose we choose the ordering M, J, A, B, E

\[ P(J|M) = P(J) \quad \text{No} \]
\[ P(A|J,M) = P(A) \quad \text{No} \]
\[ P(A|J,M) = P(A|J) \quad \text{No} \]
\[ P(B|A,J,M) = P(B) \quad \text{No} \]
\[ P(B|A,J,M) = P(B|A) \quad \text{Yes} \]
\[ P(E|B,A,J,M) = P(E|A) \quad \text{No} \]
\[ P(E|B,A,J,M) = P(E|A,B) \quad \text{Yes} \]
Notes: Belief Network Construction Details

1) order the variables (any ordering will do) A, B, C, D, E,..
2) pop the first item off the list, and test to see what a minimal set of parents is
   i) add A (trivial)
   ii) add B
      test: is it true that \( P(B|A) = P(B|\neg A) = P(B) \)
      if not, then add a link from A to B
   iii) add C
      test: is \( P(C) = P(C|A,B) = P(C|\neg A,B) = P(C|B,\neg A) = P(C|\neg A,\neg B) \)
      if so C is not linked to anything
      test: is \( P(C) = P(C|A) = P(C|\neg A) \) if so then there is no link from A to C
      test: is \( P(C) = P(C|B) = P(C|\neg B) \) if so then there is no link from B to C
      If some of the above tests fail, then there there will be links.
   iv) add D
      things keep getting uglier, but the same idea continues.
Lessons from the Example

- Network less compact: 13 numbers (compared to 10)
- Ordering of variables can make a big difference!
- Intuitions about causality useful
How to build a belief net?

• **Structure:**
  - Ask people: Try to model causality in structure
    - Real causal structures yield the best networks
  - Automatically: How to cast finding structure as search?

• **Conditional Probability Tables**
  - Ask people: Get approximate probabilities from experts
  - Automatically: Estimate probabilities from data
Belief Net Extensions

- Hidden variables
- Decision (action) variables
- Continuous distributions
- “Plate” models
  - E.g. Latent Dirichlet Allocation (LDA)
- Dynamical Belief Nets (DBNs)
  - E.g. Hidden Markov Models (HMMs)
Belief Networks and Hidden Variables: An Example
Compact Conditional Probability Tables

- CPT size grows exponentially; continuous variables have infinite CPTs!
- Solution: compactly defined canonical distributions
  - E.g., Boolean functions
  - Gaussian and other standard distributions
Summary

- **Belief nets provide a compact representation of joint probability distributions**
  - Guaranteed to be a consistent specification
  - Graphically show conditional independence
  - Can be extended to include “action” nodes

- **Networks that capture causality tend to be sparser**

- **Estimating belief nets is easy if everything is observable**
  - But requires serious search if the network structure is not known

- **Not covered: Probabilistic inference algorithms given belief nets**